DOCUMENT RESUME

ED 163 105 UD 018 786

AUTHOR Anderson, Janice K.; And Others
TITLE The U.S. Office of Education Models to Evaluate E.S.E.A. Title I: Experiences after One Year of Use.
INSTITUTION Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
PUB DATE 78
NOTE 48p.; For a related document, see ED 152 823

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Compensatory Education Programs; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Federal Programs; *Models; Program Effectiveness; *Program Evaluation; State Federal Aid; State Programs
IDENTIFIERS *Elementary Secondary Education Act Title I

ABSTRACT

In order to assess the impact of Title I programs, models specifying criteria, techniques and methodologies for evaluation were developed. These models prescribe pre- and post-test results for both Title I and non Title I children. This report describes the status of the use of the models, the experience of people who used them during the 1976-77 school year, Federal progress in drafting regulations regarding their use, and the future work planned to improve the quality of the data produced. Specifically, a brief history of the program to develop and encourage the use of models is presented. The work of the Technical Assistance Centers, the number of districts using the models, and projected use are also discussed. State and local level activities are described. A list of recommendations for refinement of materials is included. Simultaneous with the adoption of new techniques by many States and districts is the Office of Education's intent to make their use mandatory by all Title I grantees who offer services in the basic skills to school children in grades 2-12. That process, the public input, and the initial idea about content of regulations are described. Conclusions about various aspects of the Title I evaluation, reporting system, and the program to support its use are also summarized.
(Author/GC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163105

THE U.S. OFFICE OF EDUCATION MODELS TO EVALUATE E.S.E.A. TITLE I:

EXPERIENCES AFTER ONE YEAR OF USE

by

Janice K. Anderson
Richard T. Johnson
Ronald L. Fishbein
Robert M. Stonehill
Judith C. Burnes

Office of Planning, Budgeting, and Evaluation
U.S. Office of Education

1978

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UD018786

TABLE OF CONTENTS

PREFACE i

EXECUTIVE SUMMARY ii

I. INTRODUCTION 1

The Models, in brief 3

Overview of the report 4

II. STATUS OF THE IMPLEMENTATION OF SECTION 151 5

History and early years of the program 5

Number of LEA's using the new procedures 7

Expectations for use of the models next year 11

III. EXPERIENCES IN THE USE OF THE NEW SYSTEM 16

The SEA process to facilitate LEA use of the models . . . 16

The data manipulations necessary in evaluations 21

State and local staff comments regarding the models . . . 25

Suggestions regarding the new evaluation system 29

IV. REGULATIONS 37

Public participation and consultation 37

Regulation content 37

Regulations process and schedule 38

V. CONCLUSIONS 39

Conclusions and plans regarding assistance services . . . 39

Documents about the models and their procedures 39

Focused investigations to answer remaining questions . . . 40

The importance of continued public input 41

PREFACE

Amendments to Title I of the Elementary and Secondary Act in 1974 included a requirement for the Commissioner of Education to develop evaluation models and standards for use by State and local educational agencies. Experience in using the models which have been developed is the main subject of this report.

Although the models have not been required by regulation, all States have been considering the models and approximately 5500 of the school districts receiving Title I funds are now using them to some degree. All this activity advances us toward the goal of better Title I evaluations but we are especially indebted to a number of individuals for their extraordinary efforts in testing the models and sharing their experiences with us. They include Don Christie and Vince McInnis (Maine); Clindt Berndt (South Dakota); Chuck Statler (South Carolina); Arlie Cox and Carl Evans (Ohio); Oliver Himley and Ron Huff (Iowa); Mary Ann DeLong (Ocala, Florida); Mel Lucas (Gainesville, Florida); and Perry Geiger (Bronson, Florida).

As the report reveals, use of the models in school districts suggests many ways to either improve the quality of Title I evaluations or to lessen the work involved. These changes will be made as we continue to refine the Title I evaluation and reporting system.

Carl E. Wisler
Director, Elementary and Secondary
Programs Division
Office of Planning, Budgeting,
and Evaluation

EXECUTIVE SUMMARY

In August of 1974, Congress amended ESEA Title I adding to it Section 151 which requires the Commissioner of Education to develop uniform procedures--models--for use by local and State personnel in their annual evaluations of Title I projects. Also required is a program of technical assistance to help them use the procedures. Work to develop the models emphasized (1) the validity and comparability of project evaluations and (2) the flexibility necessary to make the procedures appropriate for use in most Title I sites. Prototypes were reviewed by personnel in nine States during the spring of 1975, by program and evaluation staffs in all States (and in over 400 school districts) between June 1975 and June 1976, and again by all State personnel in national training workshops later in 1976. This review process was instrumental in soliciting input from persons who would be eventually using the models and in persuading many to use the evaluation models before any formal requirement was issued.

This is a summary of an extensive report about the experiences of people using the models during the 1976-77 school year. Included in the report are descriptions (1) of the status of nationwide use of the models, (2) of State activities to change the evaluation practices they require of their local districts, (3) of local processes to collect, analyse, and report data according to the new procedures, (4) of progress, problems, and recommendations of those who are using the models, (5) of the USOE/DHEW steps to require use of the methods by program regulations, and (6) of conclusions regarding the program and its direction in the near future.

There were fifteen States in the 1976-77 school year with half or more of their school districts using the new standardized practices to evaluate their Title I projects. Those districts comprised about 25% of the total in the United States receiving Title I funds. Use of the models has grown considerably, and about 40% of the Title I districts are using them in this year (1977-78).

The increase is probably due to several factors including the substantial amount of help available to States (and at their discretion, to local districts) from ten Technical Assistance Centers established under contract to USOE in September of 1976. Their purpose is to offer advice regarding site-specific evaluation problems in Title I, to assist States in changing to the new system (helping in the installation of new computer programs, etc.), and to provide other technical support as needed by States in offering training workshops, revising evaluation materials, etc.

In the late fall of 1977, staff members from the Office of Planning, Budgeting, and Evaluation visited six of the fifteen States who had large numbers of districts using the models in 1976-77 to discuss their experiences with the system. Based on those discussions they were able to describe the 17-step, 2-3 year process the SEA administrators had pursued to change evaluation activities in their locals to correspond more closely with those advocated by USOE. The activities, including the decision to effect such a change,

identification of pilot or "test" sites, training of personnel, change in report formats, monitoring of practices, review of data, and re-iteration the subsequent year to involve more sites have been labor intensive and exhausting (at best; frustrating and problem-riddled at worst). State personnel have noted some positive outgrowths of the process, however, including greater attention to tests and their uses, more confidence in data used for State identification of good projects, and greater familiarity on the part of program personnel with evaluation goals and issues.

To use the evaluation system, LEA personnel must faithfully adhere to procedures outlined in three evaluation models concerning test selection and administration, test scoring, score conversion, data analysis, and aggregation. Conversations with staff members at both the State and local levels identified both successes and difficulties they encountered in using the techniques.

Aspects of the system which they applauded included the clarity of instructions, the thoroughness of USOE materials regarding the different approaches, the production of data which appeared more valid (with fewer results being counter-intuitive, etc.). Remaining difficulties include methods for selecting Title I students so as not to invalidate the evaluations and the conversion of test scores to values along different scales. These were identified by USOE and SEA staff perusing LEA reports. Of the 183 examined, 32 had no discernable error which was viewed as evidence that the procedures, for the most part, are feasible and usable. Work is underway bu USOE to address the remaining difficulties.

As local, State, and Federal personnel encourage each other to use the uniform evaluation models, there are also activities underway to require their use by Title I grantees. These activities emphasize public input through discussion of issues identified in a memorandum sent to all State superintendents, Title I staffs, evaluators, and interested parties and through hearings about draft rules (to be published next year, most likely). Final regulations might be in effect soon enough, then, to ensure standard evaluation activities and reporting for Title I projects the following school year. Also underway are some studies to determine when enough data will be available which are sufficiently uniform and representative to support national statements about the achievement of Title 1 students.

Plans for the evaluation program at this time and its directions in the near future include continued support and emphasis upon the services provided by Technical Assistance Center staff, greater dissemination of evaluation materials, the pursuit of further technical documentation for the models, and continued solicitation of public input. Another progress report regarding these and other aspects of the program will be available during the fall of 1978.

In sum, the Education Amendments of 1974 added Section 151 to Title I of ESEA to make grantee evaluations of that program more uniform and more usable by various administrative levels. The Commissioner of Education is required by Section 151 to develop standard evaluation models, regulate their use, and provide technical assistance to support both that use and the improvement of Title I evaluation in general. During the nearly four years since the law was passed, USOE has developed three alternative models; consulted extensively with State and local administrators; revised recommendations as appropriate; conducted national training sessions; established ten centers to provide free, on-call assistance on Title I evaluation matters: and distributed reference materials and pamphlets on specific topics of relevance (such as administering tests, interpreting scores, etc.).

Although the general reaction has been largely positive as evidenced by the voluntary use of the proposed methods by nearly 40% of the nation's Title I districts, the reality of impending program regulations requiring usage of the models has caused some objections to surface. Specifically, the Council of Chief State School Officers has commended USOE on the extent of cooperation with States for the past four years but reacts negatively to the idea of a national mandate. Although their reaction often contrasts markedly from that of other staff in their own States, we feel a need to continue soliciting input from all groups.

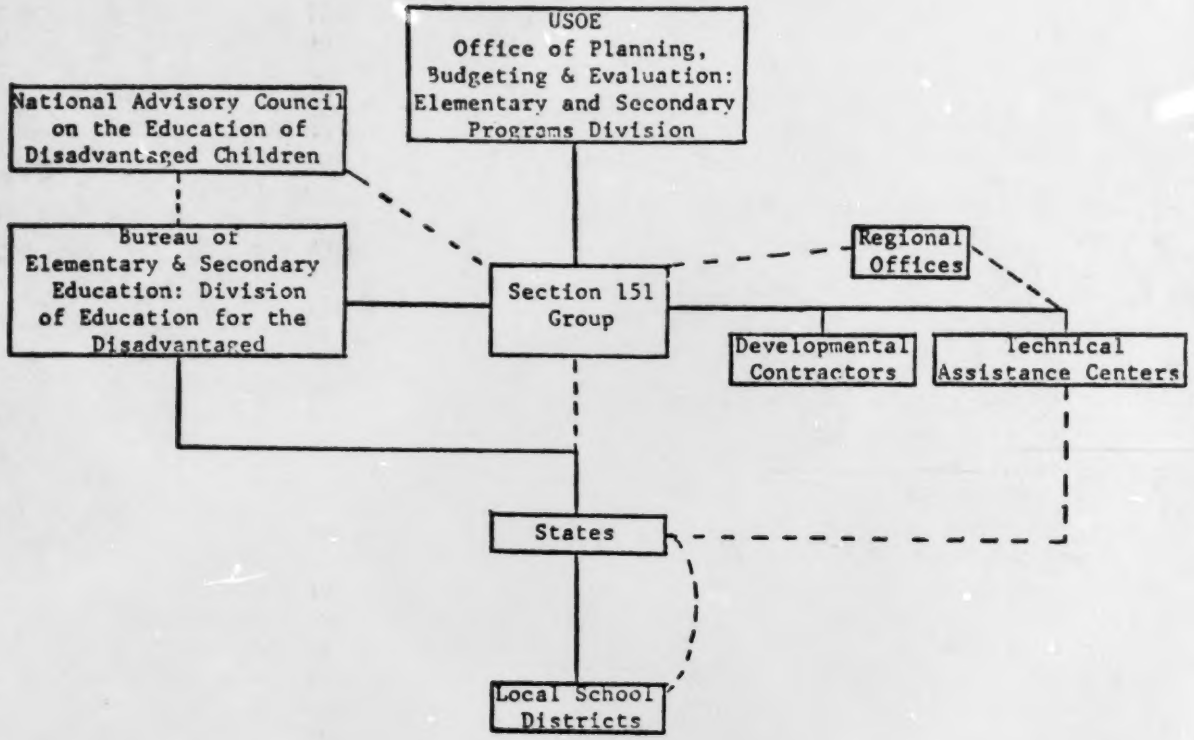
I. INTRODUCTION

The passage of Title I of the Elementary and Secondary Education Act of 1965 represented a new direction in Federal education legislation in several respects. Its goal to rectify through better schooling the disadvantagement of children from families living in poverty, its authorization for millions of dollars to school districts serving large numbers of such children, and its requirement for repeated examination of the impacts of the services funded by the program contributed to its uniqueness. In the thirteen years since its passage, these factors have continued to have an enormous impact on education in the nation as a whole.

In the last four years, especially, the evaluation component of Title I State programs has received much attention as an outgrowth of Congressional, Federal, State, and often public frustration with the quality of the data available to describe the program's impacts. Many changes have been implemented at all levels as a result of this frustration, and there seems to be reason for optimism at this time regarding data quality in the future.

The purpose of this report is to describe the status of the changes underway in districts, States, and Washington, D.C. as a result of the addition of Section 151 to Title I in 1974. Underway since the autumn of that year, the program to facilitate these changes consists of work to standardize local and State evaluation practices. (Also required as a secondary effort are national evaluations of program impacts, but they will not be discussed in this report.) Section 151 requires that the standardization of evaluation practices occur through State and local use of models developed by USOE and that technical assistance be available to them to facilitate the use of the models.

The program is administered by the Office of Planning, Budgeting, and Evaluation in the U.S. Office of Education in its Elementary and Secondary Programs Division. Also involved are personnel from the Division of Education for the Disadvantaged, the office which administers the Title I program federally. The National Advisory Council on the Education of Disadvantaged Children and personnel in the regional offices also advise OPBE regarding this effort. Figure 1.1 illustrates the relationships among the evaluation and program units, the "Section 151 group," the States, contractors, and other interested organizations.


USOE
Office of Planning, Budgeting & Evaluation: Elementary and Secondary Programs Division
National Advisory Council on the Education of Disadvantaged Children
Bureau of Elementary & Secondary Education: Division of Education for the Disadvantaged
Section 151 Group
Regional Offices
Developmental Contractors
Technical Assistance Centers
States
Local School Districts

Dotted lines signify advisory capacity; solid lines reflect communication of policy.

Figure 1.1 The various parties active in implementing Section 151 of ESEA, Title I

The models, in brief

The Commissioner of Education is required by Section 151 to develop models which "specify objective criteria," "outline techniques...and methodology," and produce data which are "comparable on a statewide and nationwide basis." The term "models," then, refers to sets of procedures regarding data collection, analysis, and interpretation which when followed will yield estimates of program impact on participating youngsters. The steps leading to such estimates are often complicated.

The perplexing question in all evaluations of social programs is to what degree a noted improvement in conditions (an increase in children's reading scores, perhaps, or a decrease in traffic fatalities, increase in voter registration, drop in number of influenza cases, etc.) is really due to the new strategy tried. Other influences on those conditions prohibit an administrator from knowing how efficacious the new strategy was, and steps must be taken to sort out from the overall change the effects of those "external" influences. The "remainder," then, is assumed to be due to the administrator's program.

The "sorting out" mentioned above is the goal of the methodology embodied in the evaluation models. Influences on children's educational performance other than the special remedial help they receive through Title I include their maturation, participation in a regular school program, greater familiarity with the tests, etc., so the steps in the evaluation models must allow the administrator to determine the degree to which Title I has helped the children over and above the improvement they show due to these other factors. The rationale behind all the models is one of observing over a period of time the progress of other children who have, presumably, been exposed to the various "other influences" on their educational performance, but have not been exposed to Title I. The difference between their progress and that of the program participants can then be reasonably attributed to the Title I services.

The three models, therefore, prescribe pre- and post-testing for the Title I children as well as for a second group. For convenience, the children in the test publisher's norming sample can act as that second group in the simplest model (Model A); a local group of children similar to those in Title I but who are not in special compensatory classes comprise the comparison group in Model B; the children in the same grade as the Title I group who scored higher on a pre-test than the Title I selection score provide the basis for comparison in Model C. If various conditions are met, statistical procedures allow the use of the data from these comparison groups for sorting out the effects of influences which rival participation in Title I to explain Title I students' improvement.

In short, to use one of the evaluation models, one must test the Title I students twice using an instrument selected because its content corresponds to the project objectives and assess the amount of student growth. Depending on the nature of the similarity between the students in the comparison group and those in Title I (as well as other factors), one uses an appropriate statistical procedure to yield an estimate of the Title I students' progress, had they not received the extra help. This estimate is subtracted from the figure reflecting their actual progress, and the difference is attributed to the Title I project.

Although certain procedures must be followed to make the estimates valid, there is much flexibility in the Title I evaluation models to allow local use of preferred tests, different scales, and most appropriate scheduling of the testing. For reporting to States and USOE, the estimates of project effects must be expressed in a metric common across sites. That is done by converting the local figures to a scale which incorporates a national distribution of scores (and for the different models, that conversion can occur at various times in the process). The national scale has been called "Normal Curve Equivalents," or more frequently, "NCE's."

Overview of this report

The remaining chapters in this report describe the status of the use of the models, the experiences of people who used them during the school year 1976-77, USOE progress in drafting regulations regarding their use, and the nature of work planned for the near future to improve the quality of the data produced.

Specifically, Chapter 2 includes a brief history of the program to develop models and encourage their use. The work of the Technical Assistance Centers, the numbers of districts using the models last year, and projections for next year are also discussed.

More detail about the use of the models last year is the focus of Chapter 3. Activities at both the State and local levels are described as well as successess and problems reported by people in those sites. A list of recommendations for further refinement of materials, etc., is included.

Simultaneous with the adoption of the new techniques by many States and districts is the process in USOE to make their use mandatory by all Title I grantees who offer services in the basic skills to school children in grades 2-12. That process, the public input incorporated into it, and the initial ideas about content of regulations are described in Chapter 4.

Chapter 5 is a summary of conclusions about various aspects of the Title I evaluation and reporting system and the program to support its use. Plans for future work in this area are also described briefly.

CHAPTER II. STATUS OF THE IMPLEMENTATION OF SECTION 151

In the spring of 1974, members of the Committee on Education and Labor of the U.S. House of Representatives decided that the Office of Education should assume a more visible leadership role in the evaluation of the effectiveness of Title I of the Elementary and Secondary Education Act of 1965. They amended the Act, adding Section 151, which defines various aspects of that increased leadership role.

Work of various types has been underway for the four years since the formulation of Section 151, and the purpose of this chapter is to describe the current status of that evaluation program. Three sections will follow: the first will summarize briefly the history of the Section 151 program; the second will describe some of its effects in terms of the numbers of LEA's nationwide using the new Title I evaluation techniques and the range of activities underway by Technical Assistance Center personnel; the third will discuss our expectations regarding use of the models next year.

History and early years of the Section 151 program

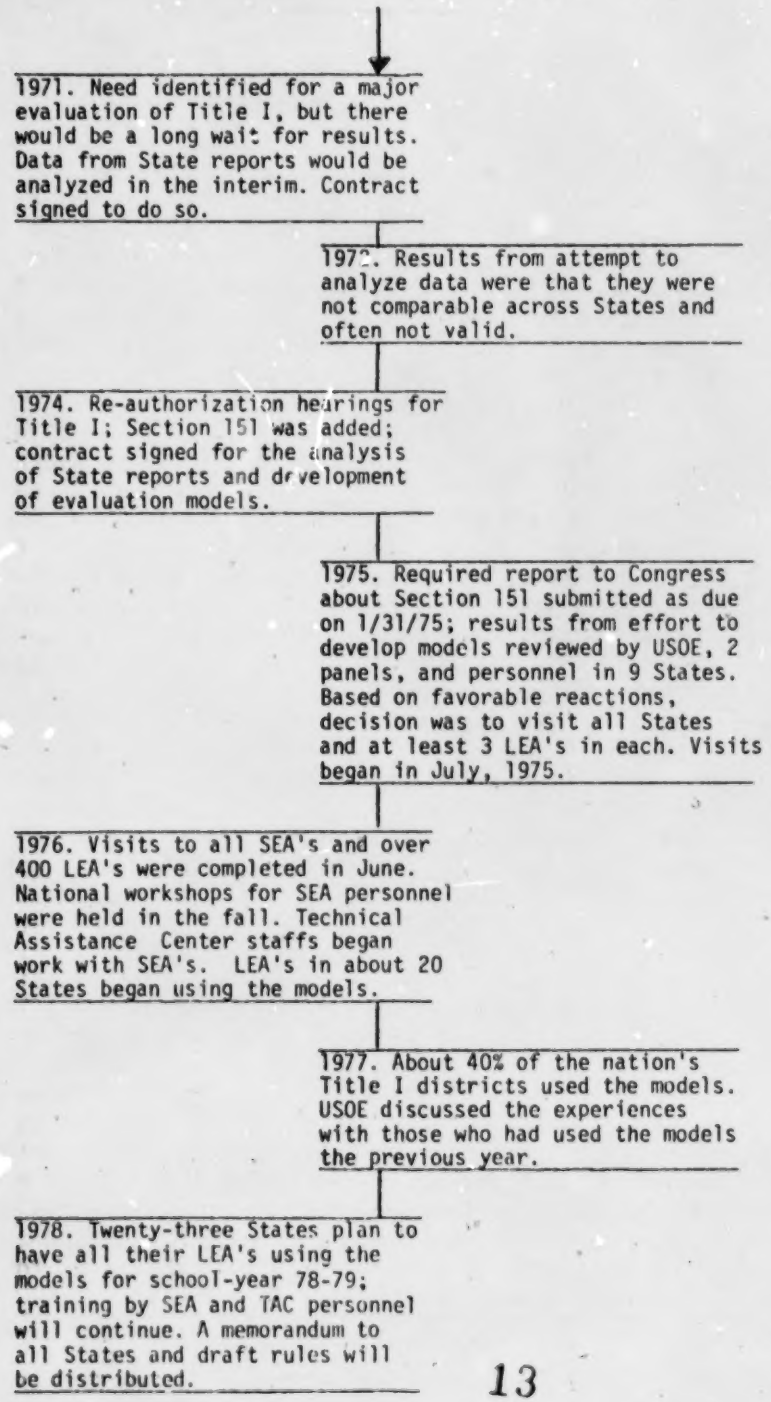
Figure 2.1 on the following page illustrates the time-line for the work in USOE to standardize the evaluation and reporting activities of ESEA Title I grantees. Early efforts to try to use SEA evaluation data were frustrated by their lack of uniformity, and suggestions were made as early as 1972 that USOE should develop methods for their use.

Reauthorization hearings in the spring of 1974 (and the subsequent passage of P.L. 93-380 adding Section 151 to Title I) strengthened the interest in those suggestions, and work was begun to develop models. Prototypes were reviewed in the spring of 1975 by USOE staff, members of two evaluation panels, and SEA staff in the states. Their reactions, in general, were that the models looked promising, and USOE initiated a new project involving visits to all States (and at least three districts of each).

The visits (each about two weeks long) continued through June of 1976, and national workshops were conducted that fall to train State personnel in the new procedures. Also in attendance at the workshops were staff of the Technical Assistance Centers established under contract to USOE in September of 1976.

The purpose of the Centers is to provide services to States; and, at their discretion to local districts, to help them address unique, site-specific Title I evaluation problems. Their work extends the more general advice provided to State personnel at national training sessions. Hence, following the workshops, the initial task of each Center was to reach a written agreement with each SEA in its region concerning the kinds of services to be provided and the best ways of providing these services.

FIGURE 2.1 Chronology of evaluation model development and use


1971. Need identified for a major evaluation of Title I, but there would be a long wait for results. Data from State reports would be analyzed in the interim. Contract signed to do so.
1972. Results from attempt to analyze data were that they were not comparable across States and often not valid.
1974. Re-authorization hearings for Title I; Section 151 was added; contract signed for the analysis of State reports and development of evaluation models.
1975. Required report to Congress about Section 151 submitted as due on 1/31/75; results from effort to develop models reviewed by USOE, 2 panels, and personnel in 9 States. Based on favorable reactions, decision was to visit all States and at least 3 LEA's in each. Visits began in July, 1975.
1976. Visits to all SEA's and over 400 LEA's were completed in June. National workshops for SEA personnel were held in the fall. Technical Assistance Center staffs began work with SEA's. LEA's in about 20 States began using the models.
1977. About 40% of the nation's Title I districts used the models. USOE discussed the experiences with those who had used the models the previous year.
1978. Twenty-three States plan to have all their LEA's using the models for school-year 78-79; training by SEA and TAC personnel will continue. A memorandum to all States and draft rules will be distributed.

In order to reach this agreement, SEA personnel had to decide which office within the SEA would be responsible for negotiating with the TAC, and which services would be emphasized. This often involved an examination of (1) areas of responsibility within the SEA; (2) SEA-LEA relations; (3) areas of technical expertise of SEA staff members and local school district talents; (4) the role of Title I evaluation in the overall goals of the SEA; (5) availability of funds for Title I and related programs; and (6) the degree to which changes, if any, would have to be made to meet the requirements of the new system. A liaison person was identified in each State as the individual through whom the TAC would clear information and activities. Next, the TAC's and State personnel negotiated a Letter of Agreement regarding the activities to be undertaken and the procedures the TAC was to use in responding to requests for assistance.

Due to the complexity of the decisions which SEA personnel had to make before a TAC could begin to provide services, reaching an agreement often took several months, even in States where the personnel were extremely receptive to the notion of technical assistance for ESEA Title I evaluation. (For those SEA's which whose initial reaction to federal technical assistance was somewhat less than enthusiastic, reaching an agreement took longer, and the provision of services often began slowly.) Thus, the TACs had little, if any, impact on the designs of the Title I evaluations used during school-year 1976-77, but, in some instances, they did provide help in interpreting results in districts which used one of the evaluation models that year.

Number of LEA's using the new evaluation procedures

Table 2.1 is a list of the percentage of districts, by State, using one of the reading and mathematics evaluation models in 1976-77. (The reader is warned that the use varies across sites with respect to quality and thoroughness As later discovered, the models were often attempted, but not completely followed. Nevertheless, the information provided in Table 2.1 can be used as a baseline regarding attempts to implement the Title I evaluation and reporting system during the 76-77 school year as well as currently. (For comparison purposes, also listed in the table is the percentage of districts in each state attempting to implement a reading and/or mathematics Title I evaluation model in school year 1977-78.)

By far, the vast majority of districts currently implementing the system are using the norm-referenced model. For the majority of Title I projects, this is the only model that will provide data that will meet the requirements of the system and provide meaningful results at the local level. Most projects do not have enough students to pursue a control (comparison) group model or a regression model. In addition, some districts have chosen to use a norm-referenced model because they feel that it is the easiest to implement (although this is not always true).

Table 2.1
Approximate % of LEA's Beginning to Implement an Evaluation Model

Region I	1976-77	1977-78
Conn - 140 Districts	25	54
Maine- 164	97	100
Mass.- 340	94	100
N.H. - 165	100	100
R.I. - 40	18	50
Vermont-54	93	100
Region 2		
N.J. - 450 Districts	0	9
N.Y. - 750	1	5
Puerto Rico - 30	100	100
Virgin Islands	0	100
Region 3		
Delaware - 23 Dist.	78	100
Maryland - 24	0	12
Penn. - 505	0	2
W. Va. - 55	9	100
D.C. - 1	100	100
Virginia - 140	0	29
Region 4		
Ala. - 127 Districts	100	100
Fla. - 67	20	35
Ga. - 187	1	15
Ky. - 181	0	2
Miss.- 152	2	30
N. Car. - 145	4	20
S. Car. - 92	100	100
Tenn. - 147	2	100
Region 5		
Ill. - 797 Districts	0	7
Ind. - 302	0	40
Mich. -575	0	0
Ohio - 625	100	100
Wisc.- 388	0	15
Minn.- 439	0	13

Region 6	1976-77	1977-78
Ark. - 385 Districts	100	100
LA. - 66 Parishes	0	10
N. Mex.-86 Districts	0	22
Okla. -613	100	100
Texas-1017	0	1
Region 7		
Iowa - 440 Districts	9	45
Kansas-250	0	100
Missouri- 389	1	70
Nebraska -300	0	100
Region 8		
Colo. - 172 Districts	0	7
Montana-250	50	100
No. Dak.- 267	0	25
So. Dak.- 188	100	100
Utah - 40	0	20
Wyoming - 42	0	30
Region 9		
Arizona - 145 Dist.	5	36
Cal. - 1080	0	1
Guam - 1	100	100
Hawaii - 7	0	100
Nevada - 17	0	100
Samoa - 1	0	100
Trust Terr.-6	0	50
Region 10		
Alaska - 48 Districts	0	19
Idaho - 105	98	98
Oregon -265	0	9
Wash. - 290	0	35

Totals Total LEAs - 13,576

1976-77 Total LEAs Beginning Implementation: 3121
Percent LEAs Beginning Implementation: 23%

1977-78 Total LEAs Beginning Implementation: 5388
Percent LEAs Beginning Implementation: 40%

In general, the regression or control (comparison) group model is most likely to be used by a medium to large project, with a large-scale testing program in place, and a computer or sophisticated hand calculator available. In some cases, where a regression or control (comparison) model is applied, the district plans also to use a norm-referenced model. (Districts have been encouraged to design their evaluations to conform to the requirements of more than one model, in case they are unable, suddenly to meet the requirements of one of the models).

The TAC's have been actively involved in providing workshops and on-site consultation to States as they implement the new evaluation procedures. As shown below, the early workshops focused on orienting LEA and SEA Title I staffs to the models. Following the orientation workshops, the usual pattern was to hold more focused workshops on particular topics for smaller groups, e.g., test selection, implementation of a model and data analysis procedures. Table 2.2 lists the percentage of LEA's in a state that have been trained; that is, those who have been presented the USOE system by TAC staff members or SEA staff members. The list below, taken from the Region VIII 15-month report, describes the workshops held and the nature of SEA and LEA consultations.

Workshops

A. Orientation--overview of reporting system and the three evaluation models

B. Model Selection--procedures and considerations when selecting a model

C. Model A Implementation--details of how to use it

D. Model C Implementation--details of how to use it

E. Test Characteristics and Selection--characteristics of test to consider when selecting one for project evaluations

F. Data Analysis--details of how to do it

Consultation

A. Pilot Site Decisions--regarding the appropriateness of LEAs serving as pilot sites for a specified model (information generated from their experiences would be used to assist other districts)

% of LEAs Receiving Some Training as of 10/15/77

REGION I (as of 12/31)	DISTRICTS	
Connecticut	140	21%
Maine	164	73
Massachusetts	340	90
New Hampshire	165	90
Rhode Island	40	45
Vermont	54	70

REGION II		
New Jersey	450	3
New York	750	0
Puerto Rico	30	0
Virgin Islands	1	100

REGION III (as of 12/31)		
Delaware	23	100
Maryland	24	20
Pennsylvania	505	80
Virginia	140	35
West Virginia	55	100
Dist. of Col.	1	100

REGION IV (as of 12/31)		
Alabama	127	100
Florida	67	100
Georgia	187	100
Kentucky	181	0
Mississippi	152	100
N. Carolina	145	100
S. Carolina	92	100
Tennessee	147	100

REGION V	DISTRICTS	
Illinois	797	37%
Indiana	302	100
Michigan	575	0
Minnesota	439	59
Ohio	625	64
Wisconsin	388	51

REGION VI		
Arkansas	385	100
Louisiana -Parishes	66	100
New Mexico	86	22
Oklahoma	613	100
Texas	1,017	5

REGION VII		
Iowa	440	77
Kansas	250	100
Missouri	389	55
Nebraska	300	75

REGION VIII		
Colorado	172	75
Montana	250	90
N. Dakota	267	90
S. Dakota	188	95
Utah	40	75
Wyoming	42	75

REGION IX		
Arizona	145	100
California	1,080	0
Guam	1	100
Hawaii	7	100
Nevada	17	100
Samoa	1	100
Trust Territories	6	100

REGION X		
Alaska	48	78
Idaho	105	98
Oregon	265	8
Washington	290	86

Totals:

Total # LEAs.....	13,575
Total # Receiving Some Training...	7,162
Total % Receiving Some Training...	52%

B. Testing Decisions--regarding characteristics of districts and of various tests which should be considered when selecting a test.

C. Data Analysis Procedures--regarding the most appropriate arrangement for data analysis (e.g., computer versus hand-analysis

D. Neglected or Delinquent Programs--meeting with SEA and LEAs to discuss evaluation procedures in the interim before models are disseminated by USOE

E. Planning Future Work--meeting with SEA to determine specific content of workshops and more general strategies for training and assisting LEA personnel

F. Reporting Forms--regarding reporting formats to meet specific needs

G. Computer Software Installation--assisting SEA and LEAs to install computer program for data analysis

Table 2.3 lists the special interests of each State and reflects the variety of the activities of Center staff. Also, copies of the 15-month report (Oct., 1976-December, 1977) from each of the Centers are available by request from the Office of Planning, Budgeting, and Evaluation.

Expectations for use of the evaluation models next year

As shown in Figure 2.1 and Table 2.1, we expect more widespread use of the new evaluation procedures next school year. There is much emphasis in SEA's this spring on revising materials and training local personnel. As of February 1 this year, all SEA's but one have a plan specifying how they will move into statewide use of the system, and in the near future, most will have achieved that goal.

Other new developments which facilitate adoption of the models are regional coordinating councils who meet to discuss common problems and solutions and the distribution of USOE's quarterly evaluation newsletter to all LEAs. Both of those activities increase the visibility of the Title I evaluation program and encourage SEA personnel to proceed with the new models.

Table 2.3
Special Interests of States

REGION I

Massachusetts

- Merging of State and Title I evaluations
- Needs assessment
- Computer/management data info systems

New Hampshire

- Process evaluation
- Use of data for decision making
- Affective and early childhood evaluation
- Longitudinal studies

Vermont

- Longitudinal effectiveness of Title I programs

Connecticut

- Early childhood evaluation
- Process data

Rhode Island

- Problems of aggregation
- Test selection; out of level testing
- Use of data for program improvement

REGION II

New Jersey

- Adapting Model C to account for local variations in achievement growth in large LEAs

REGION II (cont.)

New York

- State centralized processing of data

Puerto Rico

- Test development, norming, equating

REGION III

District of Columbia

- Evaluation design
- Development of local norms
- Use of criterion-referenced tests
- Simultaneous implementation of all three models

- Development of a laboratory handbook

Maryland

- Test selection
- Presentation and Utilization of test results
- Data collection procedures

Virginia

- Evaluation design
- Test selection
- Data collection procedures
- Utilization of test results
- Presentation of results to the public
- Overview of evaluation for Parent Advisory Councils
- Installation of computer programs
- Use of criterion-referenced tests

REGION III (Cont.)

Pennsylvania

Installation of computer programs
Review of currently available tests

Test selection

West Virginia

Evaluation design
Data collection procedures
Assisting in the development of a competency-based reading test for use in Title I
Assisting W. Virginia in developing a Title I Evaluation handbook

Delaware

Test selection & reporting forms
Reporting to parents
Inservice training in evaluation methods
Affective measurement

REGION IV

Florida

Computer analysis system
Model C

Georgia

Affective measurement

North Carolina

Affective measurement

South Carolina

Technical issues in Models

Tennessee

Computer programming

Kentucky

General orientation to systems

Mississippi

Staff training

REGION IV (cont.)

Alabama

Early childhood evaluation
Functional level testing

REGION V

Illinois

COBOL computer program
More detailed test information; relate evaluation to state assessments

Indiana

Include grades 1 and 2 in evaluation

Michigan

Reduce norm referenced data collection burden for districts using ORTS. Coordinate evaluation with state assessment.

Minnesota

Install computer program on Burroughs.

Ohio

Include grades 1 and 2 in evaluation.

Wisconsin

More detailed test information.
Process evaluation

REGION VI

Arkansas

Test administration
Test score interpretation
Model A1 implementation
Migrant evaluation

REGION VI (Cont.)

Louisiana

- Non-normed testing
- Awareness of PACs
- Computer support
- Hand calculators
- Objective writing

New Mexico

- Reporting forms
- Model AI implementation
- Affective measures
- Regression models

Oklahoma

- Computer support
- Data analysis

Texas

- All models implementation
- Migrant evaluation
- Essential of good process evaluation; Neglected and/or delinquent models

REGION VII

Iowa

- Installation of computer system;
- Special study on quality control

Nebraska

- Installation of computer system;
- Special study on base-line data

Kansas

- Computer program documentation

Missouri

- General orientation to system

REGION VIII

No. Dakota

- Early childhood evaluation

So. Dakota

- N/D evaluation

Wyoming

- Using affective measures
- Use of Title I evaluation for local decision making
- Integrating Title I Evaluations with State data collection
- Test selection issues
- Process evaluation

Montana

- Use of CRTs and local norms

Colorado

- N/D evaluation and CRTs
- Migrant evaluation
- Process evaluation
- Coordinate Title I and other testing

Utah

- Computer System
- Evaluation involving minority groups; esp. Native American
- CRTs and local norms

REGION IX

Arizona

- Local norms
- Regression problems (Model C)
- Combining NRTS with non NRTS
- Upgrading staff skills

California

- Model C and A pilots
- Combining Calif. assessment with Title I

REGION IX (cont.)

Hawaii

Local use of Title I results
Process evaluation
Test selection

Nevada

Modification of computer output to get individual scores

Guam

Instrumentation problem on TESOL
Selection and development of instruments on TESOL
Improving selection & placement of children

Trust Territory

Developing own tests and local norms

Samoa

Develop plans to coordinate Title I and other Federal programs eval's
Improve evaluation capabilities

REGION X

Alaska

Developing local norms

Idaho

Improving quality of Title I eval.

Oregon

Establishing computer network
Use of Rasch Scaling in Title I evaluation.

Washington

Improve staff evaluation capabilities
Expanding utility of Title I computer programs for feedback to teachers

CHAPTER III. EXPERIENCES IN THE USE OF THE NEW EVALUATION SYSTEM

The purpose of this chapter is to describe (1) the experiences of State and local personnel who used the evaluation models to assess project effects for the 76-77 school year, (2) the successes and difficulties they experienced, and (3) various suggestions (from them as well as from others) for making the procedures less difficult. To put their experiences in context, we will first outline the steps necessary at the State level to get the models used by LEA's. Secondly, the data manipulation prescribed by the models, themselves, will be described such that users' suggestions for refinement can be related to the overall process.

Hence, this chapter has four sections: (1) the SEA process to facilitate LEA use of the models; (2) the data manipulation necessary in project evaluations; (3) the experiences of personnel at the SEA and LEA levels in administering or performing the evaluations; and (4) suggestions for modifications in the USOE - recommended procedures.

The SEA process to facilitate LEA use of proposed USOE evaluation models

During the 1976-77 school year, 28 of the SEA's across the nation had LEA's using one or more of the proposed models to evaluate the reading and/or math outcomes attributable to participation in a Title I project. (See Table 2.1) for the most part, those sites used the techniques for the first time. Information gleaned from the evaluation personnel there gives an idea of the administrative procedures involved in effecting the change to new methods.

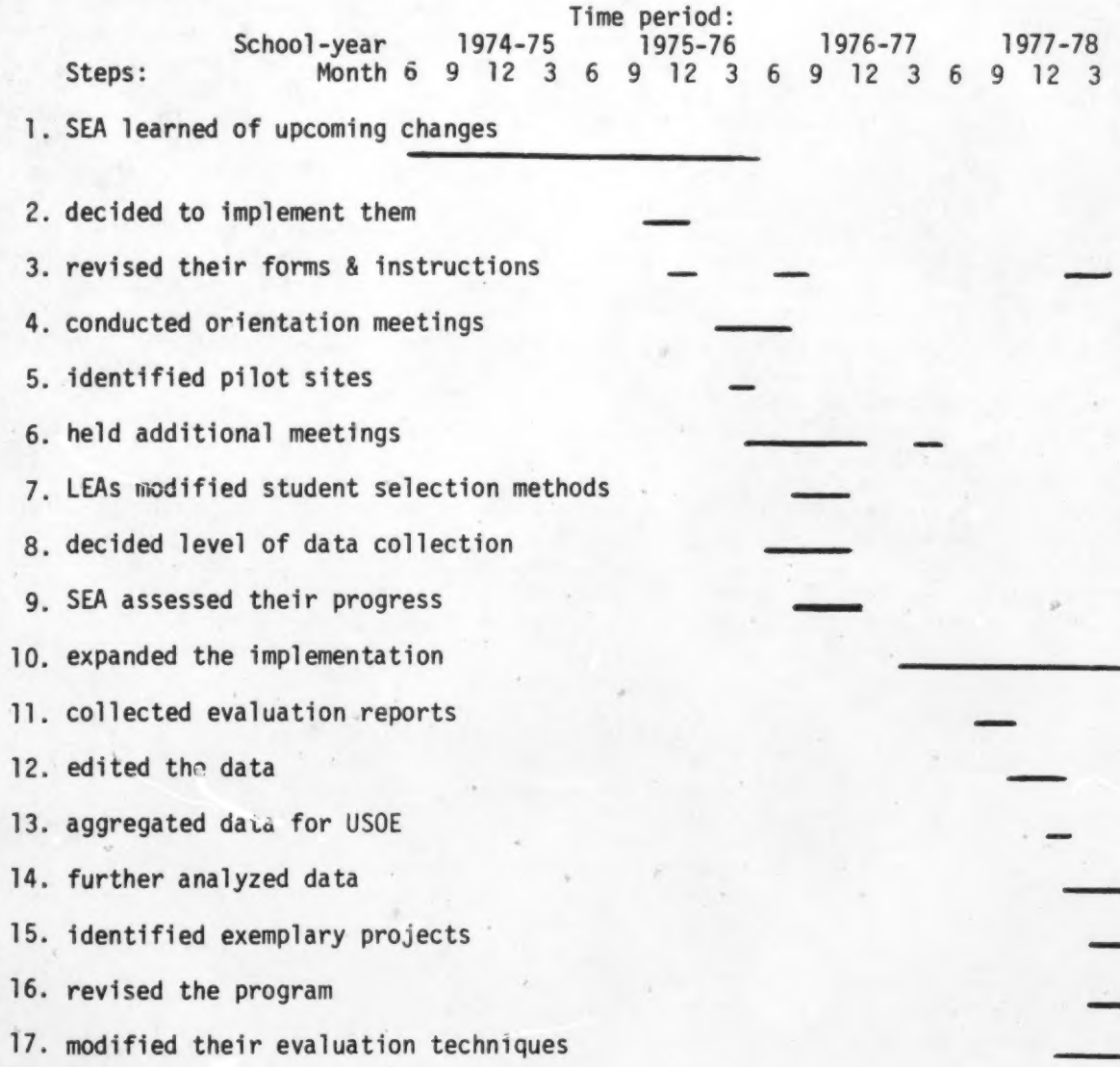
OPBE personnel contacted six of the SEA Title I directors who have had the longest and most comprehensive experience in using the new procedures. Without exception they were willing to share information concerning the characteristics of the implementation process. Visits were made to the states of Florida, Iowa, Maine, Ohio, South Carolina, and South Dakota.

The generalized step-by-step procedure followed by a SEA is presented below, and a GANTT chart (Figure 3.1) illustrates typical timing of those steps. The procedure does not represent a specific state, but rather a composite of what happened in the States whose staffs we contacted.

1. SEA and some LEA personnel in every state learned of upcoming changes through USOE presentations at national meetings, visits to the states by RMC Research Corporation personnel (under contract to USOE), and through visits by various USOE staff members. For SEA personnel, this "first exposure" to some of the new ideas could have been as early as summer of 1974 (USOE was already recommending the avoidance of certain test score metrics in conjunc-

Figure 3.1

Sequence and timing of the implementation steps in a typical state.


Time period:
School-year 1974-75 1975-76 1976-77 1977-78
Steps: Month 6 9 12 3 6 9 12 3 6 9 12 3 6 9 12 3
1. SEA learned of upcoming changes
2. decided to implement them
3. revised their forms & instructions
4. conducted orientation meetings
5. identified pilot sites
6. held additional meetings
7. LEAs modified student selection methods
8. decided level of data collection
9. SEA assessed their progress
10. expanded the implementation
11. collected evaluation reports
12. edited the data
13. aggregated data for USOE
14. further analyzed data
15. identified exemplary projects
16. revised the program
17. modified their evaluation techniques

tion with the effort to document and disseminate information about effective practices), or as recently as the 1975-76 school year.

2. In some cases, SEA staff decided to pilot, field test, or fully implement the new system, occasionally with limitations on their LEA's as to choice of model or test. Evaluation and program personnel discussed the ramifications of adopting the proposed models, balancing technical and political considerations against the expected return of useful information. In some States, these discussions disclosed a preference for Statewide use of a single model. In others, a decision was made to encourage the LEA's to consider all the models, and choose the one(s) best suited to district characteristics.

3. Following the SEA decision to adopt the model(s), staff members revised their evaluation reporting forms used for collecting data from LEA's. This revision was based on the types and extent of information required for implementation and for quality control. Generally they attempted to reduce the paperwork formerly required from the LEA; starting with the draft forms produced under contract for USOE and sent to SEA's for review, they added sections dealing with specific material needed by the SEA (such as data concerning the nature of the project, whether it was pull-out or an intact class, if aides were used, overall program emphasis, etc.) Even with these additions, however, the result was an abbreviated reporting form compared to those from prior years.

4. The State then conducted orientation meetings to publicize the change of procedures, to train local personnel regarding the new forms, and to obtain information on problems which could arise. Material covered at the meetings included an orientation to the evaluation models, discussion of the proposed LEA report forms, and notification regarding the implementation timetable. State personnel stressed the reduction of paperwork (and possibly also of testing) as a compensation to the LEA's for the increased rigor demanded in evaluation.

At all the workshops, attempts were made to keep the number of participants to a minimum (to allow interaction among them and to allow attention to each person). This meant that since attendance of at least one person from each LEA was compulsory, States had to hold several workshops over a period of weeks. Often the sessions ended with the SEA distributing a questionnaire regarding the attendees' reactions to the changes, thoughts about other types of workshops or assistance they might require, and probable choice of model.

5. Usually SEA's identified pilot sites to try the models the first year (75-76). Since the workshops generally took place after the selection of Title I participants and their pretesting in the schools that year, the full-fledged system could not be started immediately. Consequently the SEA chose a subset of the schools in the State to pilot the models, and to use the new forms and instructions if they desired. The data collected from the pilot school districts the first year were generally somewhat flawed for project assessment since most of the LEA's had problems following all of the procedures for any given model. The experience was considered valuable, however, as a basis for determining remaining difficulties, most troublesome requirements, and further needs for technical assistance.

6. Coincident with the process of choosing try-out schools, the SEA held additional meetings. Technical materials were distributed to all pilot LEA's, and their participation requested at area workshops. SEA personnel, by this time much more conversant with the new procedures, gave detailed instructions on the model requirements. They were often assisted by Technical Assistance Center personnel (as described in Chapter II). The State leaders helped LEA evaluators try out actual analyses on sample data. Those States in which the SEA program personnel played a major, visible role in the workshops found that their early involvement increased the chances of success.

7. With SEA encouragement both in the meetings referred to above and in frequent mailings of technical papers to the schools, the LEA evaluators modified their methods for selecting children for Title I services. Since most of the LEAs chose to use Model A (the norm - referenced approach), they were required to select Title I participants on something other than their pretest scores. (Previously almost every LEA had selected students on the pretest, thus leading to estimates of the efficacy of Title I artificially inflated due to the regression effect.) Suggested new methods included the use of diagnostic tests, failing grades in the classroom, objective teacher-rating scales, district-developed tests, and subscores of standardized tests.

8. The LEAs also decided on the most appropriate level of data collection for their situation. The evaluation system requires that data be collected on a "project;" hence, any information obtained should describe that unit. If a project were defined on a "micro" level such a s a single language arts classroom, the number of students comprising the observations about that unit might be too few for stable conclusions; if defined on a "macro" level (for example, including all language arts instruction in a district), perhaps too little detail would be available for use in identifying the educational practice responsible for the change. Thus, a project might include classrooms which varied on hours per week instruction if the LEA evaluator considered the time difference to be inconsequential. But, for instance, data from classrooms in which computer-assisted instruction was used would not be combined with those in which peer tutoring was the mode because the two instructional approaches are qualitatively different.

9. At the conclusion of the first year of implementation, the SEA assessed its progress thus far:

a. Had all evaluation personnel in the State been at least exposed to the new system?

b. Had a pilot group of LEA's attempted to use the new methods and report forms?

c. Were the reporting forms being revised and becoming simpler to use?

d. Had LEA procedures for selecting Title I students been modified, as appropriate, to avoid the contamination of the evaluation data?

e. Were non-pilot-testing LEA's accepting the models (no longer hoping the change would simply go away)?

f. Was there an awareness among the State personnel of the need to establish some quality control procedures?

10. In the second school year, 1976-77, the State expanded the implementation. All schools in the State were required to use the models and new reporting forms. Additional workshops were provided for those who wished to review the material, and for new personnel recently given responsibility for Title I evaluation. Often, a second workshop during the school year was added for discussing the mechanics of completing and interpreting evaluation reports.

11. At the conclusion of the first complete school-year cycle, the SEA collected evaluation reports from all LEA's. Personnel in the State office scanned the reports, and if they appeared to be complete, they were processed. Otherwise they were sent back to the LEA's for correction.

In the States with separate program and evaluation units, at least two copies of each LEA report were collected. One copy was given to the program office and the other to the evaluation staff. A serious attempt was made by the State Title I director to involve the entire SEA staff in evaluation, and to use the results in modification of the program.

12. As a first task in the analysis, the SEA staff edited the data to determine its quality. The general characteristics of the project as reported were matched with what had been proposed in the project's funding application. Often data from entire grades were missing, or the numbers of students reported differed greatly from the number anticipated. Appropriateness of tests, forms, and levels, and testing dates were ascertained. Telephone calls were made to LEA's with incomplete or erroneous returns; the worst reports were returned to the schools to be redone.

Then actual data were checked. This included inspection for out-of-range values, excessive variability across projects or grades, a large proportion of scores in the guessing range, incorrect arithmetic, unusual changes of number of students across grades, etc. Often, the SEA found that much better editing was possible when individual student scores were reported to the State, so they modified their forms to request that LEA's cut students' names off the worksheet, and send the scores along with the report.

Another part of the data quality control was the checking of results from sites where evaluation was known to be well done in order to estimate the range of project gains likely to be found in the State. Often such superior sites were large cities where the variability across projects was great. Hence, such data were felt to establish reasonable bounds for project estimates; if data from projects fell outside that range, they were re-examined for possible errors.

13. The SEA personnel aggregated the most error-free data for their report to USOE. Copies of the State report were sent to USOE, to all LEA's within the state, to other SEA's, and to representatives in the State and federal legislatures. Sometimes the reports were given to the newspapers.

14. After the required report to USOE was completed, the SEA further analyzed the data to answer State-specific questions. For example, gains were calculated for projects aggregated by the tests used for evaluation in order to determine whether certain tests gave an evaluation "bonus" to the user. Or gains were calculated for projects involving pull-out programs and compared with those involving aides.

15. The evaluators within the State also identified exemplary projects. Those with large consistent gains across grades were selected, and if the appeared to be reliable and valid, the results were disseminated. Projects with large consistent losses were also examined, and the State program personnel made suggestions about possible methods of improvement to the LEA's.

16. Not only the extremes were investigated; based on the complete evaluation results, State personnel considered revisions to aspects of the program to reflect what had been learned. Where they were certain that greater gains were possible in the lower grades, for instance, they encouraged LEA's to emphasize instruction at that level. If the use of peer tutoring resulted in generally larger gains, they stressed that possibility to LEA's for the next year's applications.

17. And last of all, the SEA modified the evaluation procedures to improve the quality of the data and increase the scope of the decisions that could be guided by those data. Many of the specific problems they addressed are delineated in the following section, and possible remedies are suggested.

The data manipulations necessary in project evaluations

There exists tremendous variety, both at state and at local levels, in the ways that Title I evaluation information is gathered, analyzed, aggregated and reported. In the extremes, some States currently gather data at the individual student level from the LEA's and then employ central processing while other States rely entirely on their LEA's to analyze their own data. Some States have automated virtually the entire evaluation process using scoring services and automatic data processing; others rely on information that is scored and analyzed entirely by hand. In States that have some very large and some very small LEA's, differences in procedures will vary tremendously even within the State.

Any explanation for this wide variety of practice across the national Title I system would consist of many factors -- capabilities and support systems vary widely, different philosophies and priorities are used to set policies, the numbers of Title I students in a district may range from the tens to the tens of thousands, while the amounts of Title I grants might range from the tens of thousands to the tens of millions of dollars. Thus, it is difficult to envision, at the present time, a unique correct way to do things; the system must be designed to function efficiently with alternative methods for data handling, processing, analyzing and aggregating.

A schematic diagram of achievement data flow through the Title I evaluation system is presented in Figure 3.2. It is arranged to display the tasks typically done at the LEA and the SEA level. Depending on the sophistication of a particular SEA or LEA, a task (such as conversion of scores) can be undertaken at the SEA, LEA or both levels. Where a line in the chart crosses to the SEA portion of the figure, that represents data being reported by the LEA to the SEA, in the form indicated by the parallelogram where the line originates. The roman numerals denote phases to be discussed below, including test selection and administration, scoring, score conversion, data analysis, and data aggregation.

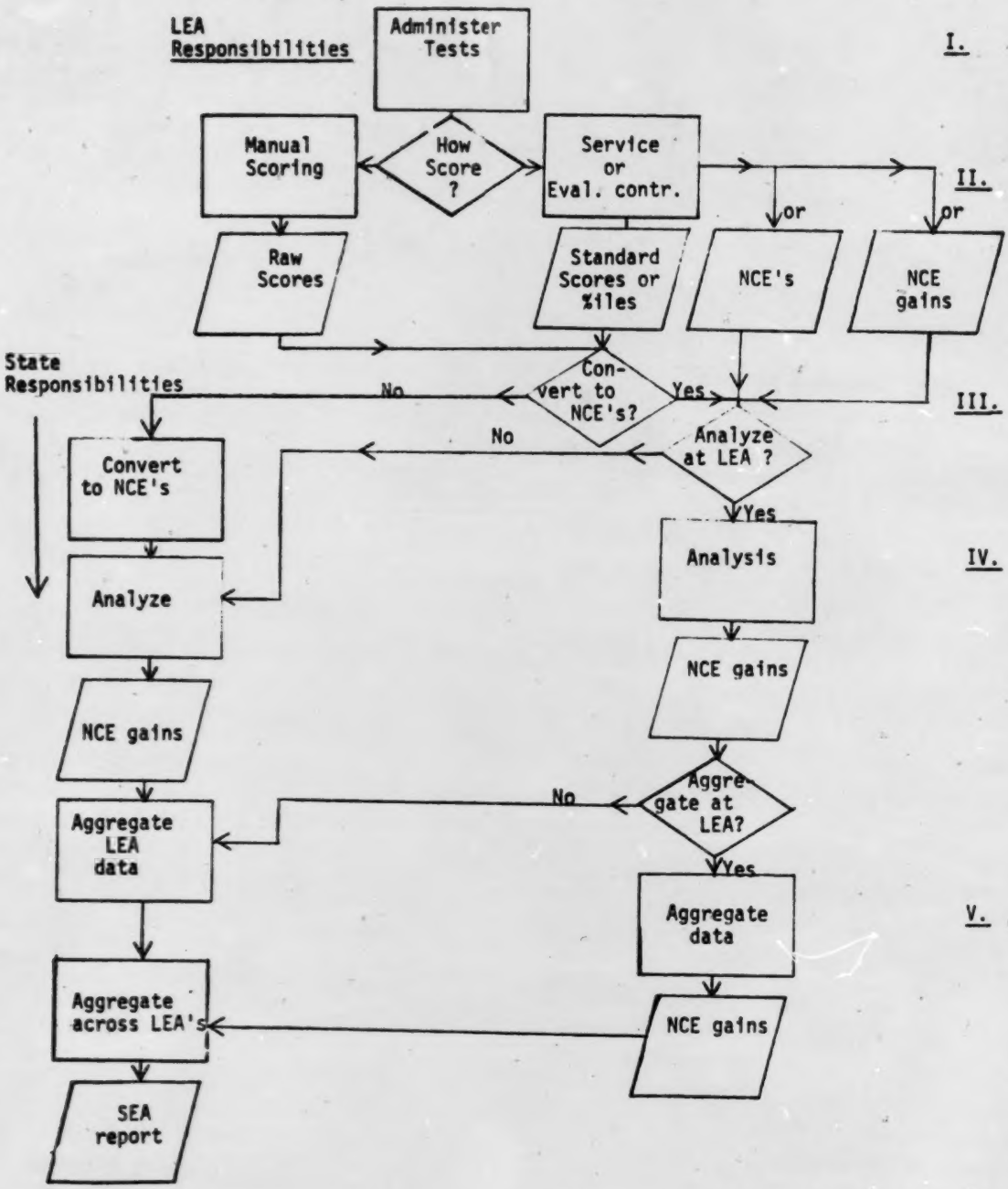
Phase I - selection/administration of tests. At this beginning phase of an evaluation, several steps are important: (1) the proper test must be selected; (2) the correct level of it must be administered; (3) the test administeration procedures must be standard; and (4) the testing conditions must be appropriate. There are many aspects to each of these principles, but they can be summarized as follows.

The "proper" test is, foremost, one which measures what is being taught. Much research (Armbruster, Stevens, and Rosenshine, 1977; Bianchini, 1976; Hoepfner, 1976; Stearns, 1977; and Tallmadge, 1977) has highlighted the degree to which different standardized tests emphasize different subskills within a skill area. Title I evaluators are advised that the more closely their tests correspond to the project objectives, the more relevant the scores will be for detecting student growth in the project (Fagan and Horst, 1978). This fairly common-sense notion is often disregarded as other factors influence test selection.

Furthermore, the "proper" test has empirical normative data on children similar to those in the project at times of the year corresponding to the pre-and post-test dates for the evaluation (if Model A is being used). Also, the norm data are from a representative national (or local) sample of children.

Using the correct level of the test means administering one on which the fewest children possible score either at the "chance" level or at the top score. This is important because a preponderance of the former (many students at the "floor" of the test) artifically inflates the group's pretest average, thereby overstating their status before the project. An estimate of their gain due to the project would of course then be underestimated. (Similarly, if they "top out" on the post-test, the group's status after the project is underestimated, and the resulting gain figure is again too small.) Of course,

Figure 3.2 Flow of Data in the ESEA Title I Evaluation and Reporting System



the mismatch of test level with student skill levels can also affect evaluation results in other ways; the important consideration is that students' performance levels be reflected as accurately as possible. Use of the wrong level of the test precludes this (Roberts, A.O.H. 1978; Roberts, S., 1978).

All tests have standardized procedures outlined for their use. (Even "home-made" instruments have instructions for administering them.) Such procedures may include timing, use of practice items, degree of assistance from the test proctors, etc. In order that students' test scores be comparable to those of others (especially to the norm data), these procedures must be followed (Horst, 1978).

A related matter is the insistence on good testing conditions. These include quiet rooms, similar settings for testing project and "comparison" group children, same time of day, etc. (Horst, 1978; Tallmadge and Roberts, 1978).

Phase II - scoring of instruments. This step involves determining the test score for each student. Occasionally it is done by hand: teachers or clerical personnel place a grid over each answer sheet, mark the incorrect responses, and tally the student's score. More often, a service is used. Student answer sheets are sent away to a firm who will return lists of students tested, the raw score for each, and requested score conversions. Problems may arise with lost answer sheets, incorrectly coded ones, damaged pages, etc.

Phase III - conversion of scores. Due to the mathematics required by the analytical procedures in an evaluation, students' raw scores (number right) must be converted to a standard score metric. In most cases, the preferable standard score metric for the computations is one which incorporates characteristics of the national distribution of scores for the age group--the NCE. In order to derive that figure, as many as three or four separate conversions may be necessary for each student's score. (A typical sequence would be the child's raw score, to the publisher's standard score, to a national rank or percentile, to an NCE).

Some scoring services can provide all of these scores; all can provide the percentile equivalents. There is a charge for each additional score requested, however, so the use of a service for all of the conversions could be expensive for large projects. The trade-off is one of accuracy, it appears. (See next section regarding comments from people about their experiences.)

Phase IV - data analysis. This is the phase in the Title I evaluation where data from individuals are combined into project-level statistics. NCE gains, describing the effectiveness of the Title I project in contributing to the students' learning above and beyond what is expected from the "regular" curriculum, are calculated. Each model prescribes the appropriate analytical techniques, which range from fairly straight-forward computations in the case of Model A to complex statistical manipulations in Models B and C.

Phase V - data aggregation. This is the final step in the Title I evaluation system, prior to the actual reporting of the evaluation results. Errors made at this stage of the evaluation are not likely to be too serious, since the data on which the aggregations are based (project-level NCE gains) are usually accessible for later double-checking. It may take place occasionally in a school district central office (when a single project is implemented in several buildings) but more often this phase occurs at the SEA level.

The first two sections of this chapter have described the implementation of the evaluation models-- in terms of both (1) State administrative efforts to bring about a change in LEA activities and (2) those activities, themselves. The third section will describe the first experiences with the models of both groups.

It is important in this discussion to recall the purpose of this "field test" and documentation effort. Under Section 151 of ESEA, Title I, the Commissioner of Education must require SEA's and LEA's to use "techniques... and methodology...for producing data which are comparable on a statewide and nationwide basis" (Subsection f). The development and publication of valid evaluation models as required also by Section 151 are the first steps toward the nationwide production of comparable data; later necessary steps are the monitoring of the use of those models and the development of means to make their use more error-free.

Hence, we must re-emphasize that the purpose of this investigation is to assess to what degree the evaluation system as now defined (three evaluation models, aggregation guidelines, technical documentation, a computer program for producing reports, etc.) can yield the comparable, uniform data required by the legislation. We will also address the benefits States and locals have attributed to the use of the models; and where we find that the use of the system has not gone as intended, we can pursue means to improve the system. Hence, the following section describes the findings of conversations with various system users; the final section lists some suggestions from them for refinements.

State and local staff comments regarding their experiences with the models

Both State and local district personnel have shown much progress in using the models and note already some positive outcomes attributable to that progress. One thing especially welcomed by State personnel is the provision of comparable data across districts. (Many have never had data reported from districts which could be aggregated or compared.) Most States plan to--and some already do--use the evaluation data to identify especially effective strategies within the State.

Another favorable outcome has been the existence of standards which States can use for advising and monitoring their LEAs' evaluation activities. Some feel that the lack of such information historically left them with little basis for insisting upon specific LEA evaluation practices.

State personnel have also noted the benefits of greater attention to achievement tests--their content, use, selection, etc. In many cases, testing practices have been favorably influenced by the models' recommendations of certain procedures depending upon test characteristics. That has prompted everyone to look more deeply into information about their tests.

Local personnel share to a large part the positive reactions to learning more about tests. They also welcome the possibility of comparing the outcomes of their efforts to those of districts they know to be similar. In many cases there is also a positive reaction to the explicit statement of evaluation steps. They feel the recommended procedures are less a burden than their former attempts to meet unclear mandates.

Since the system does establish standards and recommendations for both groups, it also enables examination of the extent to which whose standards are being met. In a sense then, one must acknowledge that any analysis of evaluation practices will yield less glowing comments than were prevalent 8-10 years ago because at the earlier time there was no way to judge anything as unacceptable. We solicited the opinions of our program and evaluation contacts regarding their problems, recognizing that the definition of "problems" is a relative one. We suggested they determine problems in comparison to an ideal implementation of the evaluation system, realizing that relative to former practices, they often represent considerable improvements.

An example of this sort of dilemma with a problem that by its very existence marks improvement over former practices is the first listed under Data aggregation in Table 3.1. The Title I evaluation system emphasizes the aggregation of data only for projects with evaluation data on a major portion of the children served (projects with minimal attrition). This is to ensure that the evaluation data truly reflect the progress of project participants; and before this can even become problematic, one has to have progressed from common earlier practices of tabulating pretests and posttests and never checking that they were for the same children (allowing, in the worst case, the assessment of a group's posttest compared to a preceding "pretest" of completely different children). Hence, in many cases, the list in Table 3.1 is evidence of an improvement in local evaluation methods.

There remain areas for further improvement, however, and they are discussed in three categories: procedural, relating to adhering to suggested rules of the models; clerical, referring to recording, translating, and calculating; and analytical, resulting from technical or statistical problems.

Procedural. One area still needing work relates to the general implementation rules. The requirement of Model A that selection of Title I participants be based on data other than from the pretest suffers the most abuse. Even among the States which were the first to attempt the models (States with extremely competent Title I personnel), the proportion of LEA's still selecting on the pretest exceeded half of those reporting.

SEA personnel were beset with questions from the LEA's on possible methods of fulfilling the model's rules without deviating from their present practices or perceived requirements. Various personnel advised their LEA's of valid ways to choose students and still not invalidate the evaluations, but seldom were the alternative methods tried; or when they were tried, they did not eliminate the problem completely. For example, one evaluator provided tables to help teachers in student selection suggesting a choice based on a combination of standardized achievement test scores, absenteeism, a teacher's estimate of the student's achievement, an estimate of underachievement, an estimate of motivation, and health. When the reports were returned, most school districts were found to have used the pretest standardized test score combined with an estimate of underachievement (obtained by the subtraction of the student's score from the class average). Thus only a single criterion was used and the choice of individuals to be placed in the Title I group was based solely on the pretest.

The second most common lingering procedural problem is the administration of norm-referenced tests at the proper time in Model A. Comparisons between the Title I group and the test publisher's norms are most valid when based on real data points. Consequently the tests should be given during a four-or six-week period spanning the test publisher's main norming date. This was seldom the case, although at least the discrepancy between true norming date and date of use is decreasing.

Another administrative problem is that of communicating what the new NCE metric means. According to a number of State leaders, parents accept the idea with little hesitation, but administrators (especially superintendents) resist beyond belief. The use of a variety of metrics for sharing results with interested parties will help in this area.

Administrators also question the source of funds for the testing of students in Title I evaluation. The requirements of Model C include the testing of comparison students from among those not in the project. Therefore if the district already has budgeted funds for testing once a year, and selection of Title I students on that test is acceptable, then Model C is a logical choice. But if the money must be located elsewhere, the choice can just as logically be Model A. Any modifications of the regulations concerning who pays for testing would have a direct effect on choice of model and frequency of testing--perhaps more often a move to Model C with testing once a year. (This involves a small amount of money usually.)

Clerical. The translation of a raw score into any other score is fraught with error. In one State, data from more than 93% of the LEA's contained at least one table look-up error. Another State director reported that the majority of errors in his workshop exercises stemmed from the inappropriate use of the same norm table for both pretest and posttest score conversions in spite of very obvious table titles. Where yet another had modified the format of the publishers' norms tables, the error rates were considerably reduced--but still excessive. Another approach to improving clerical accuracy was the re-arrangement of the recording sheet for students' pre-and post-test scores.

When the LEA eval[illegible]nverts those scores to NCE's, calculates a different score for e[illegible]nd obtains an average difference (gain) score for the group, four [illegible] for error exist. (The raw or percentile score conversion could b[illegible] the pretest, or the posttest, or the subtraction or the calcula[illegible]uld be wrong. The subtraction of pre-post scores, and [illegible] are often double-checked, but the score conversions [illegible] provement in quality control would likely ensue from a [illegible]he worksheet as well as that of the conversion table.

Another prob[illegible] is that often the gain score for an individual turns out negative, and negative numbers appear to be an anathema to proper calculations. One evaluator was so upset by negative scores that he ignored every one in averaging gains in his project.

The assignment of a pretest or posttest score to every individual is only part of the process--the two tests must be matched for each individual. In large districts the matching is generally done on a computer. Here several difficulties appear: a matching program must be written; bad coding or punching takes a toll of properly matched individuals; a single unmatched card in a file makes the entire remaining ones mismatched and makes the evaluation worthless if left undiscovered; students change the spelling of their names; and inconsistencies appear in using the last name first.

In smaller districts, the match is carried out by hand. Here the human intelligence can remedy many of the problems: one can see that "Bill Thorn" on the pretest is "Thorne, W." on the post, for example. A knowledgeable individual might remember that a school boundary was recently changed, and half the students are in the neighboring school, or that the Jones children oscillate regularly between two schools. Thus the increased difficulty in matching by hand is offset by the increased frequency of matches. Sometimes the individual cannot find a match between pretest and posttest and simply leaves the space blank. Of the 182 forms seen by USOE staff, 29 were incomplete; the person either overlooked or ignored many items.

A comparatively minor error in reporting is the failure to include the day of the month on which testing was accomplished. Under Model A a user of the ITBS would be expected to test within a two-week interval on either side of April 28th. If he merely reported that the test was administered in April, the State evaluator cannot assume that the test was given at the proper time; it could have been on the first of April. A similar argument holds if May were reported. The error is minor in that it can be corrected easily with a change in the reporting requirements.

Analytical. A variety of technical questions dealing with the statistical and psychometric aspects of the system continue to plague the evaluators. For instance, are they penalizing themselves when they test once a year in the spring? If students forget a great deal during the summer, perhaps they would show greater gains if they were tested fall and spring. When students repeat a grade, what pretest score should be used--the first pretest or the second? And what norms should be used?

Some State personnel note that the correction of an LEA's evaluation error may result in a lowered gain estimate. (They recognize, however, that no one wants to be fooled into assuming successes in remediating children's educational problems if that's not the case.)

The sample of individual project reports perused in the six States visited by OE staff were rated on their quality of evaluation. Those projects with the "best" evaluation (as far as we could tell) showed low but positive gain scores for the Title I group. (The correlation between evaluation quality and size of gain was - .25) Those projects which appeared to be evaluated correctly showed a modest positive impact from the program.

When the directors and evaluators in the SEA's and the LEA's identified the problems outlined above, they also suggested a variety of solutions. Some approaches they had already tried out, and other possibilities arose in discussions with USOE staff. The suggestions are described in the next section.

Suggestions regarding the new evaluation system

People we visited and other interested parties have suggested ways USOE can help SEA's and LEA's follow more completely the procedures outlined in the evaluation and reporting system. Their suggestions are summarized below according to the same categories used in preceding sections: procedural, clerical, analytical.

Table 3.1 Error rates in evaluation reports

Of the 182 project reports reviewed, 32 had no discernable error; however, the remaining 150 appear to need further improvement in one or more of the following:

	Criterion Not met
I. Test selection/administration	
a. use apparently appropriate test or subtest (content-wise)	3
b. use test with relevant norm group	?
c. pretest at true norm time	66
d. posttest at true norm time	58
e. if slightly miss pretest norm date, make posttest error the same	1
f. use same level and form pre-and post with Model A	14
g. give make-up tests no more than 2 weeks after original	?
h. give low enough level of test to avoid floor effects	3
i. give high enough level of test to avoid ceiling	1
j. give appropriate out-of-level test	3
k. use test administrators who are objective	?
l. follow standard instructions, timing	?
II. Scoring	
a. check for accuracy of individual student scores	1
b. make sure student/score match is correct	?
III. Conversion of scores	
a. use correct conversion tables	11
b. do conversions accurately	29
c. interpolate between tabled values when necessary	3

	Criterion Not met
IV. Data analysis	
a. use data exclusively from students with both pre-and post-test scores	?
b. do all arithmetic accurately	32
V. Data aggregation	
a. use data from projects with only minimal attrition	5
b. check that drop-outs (attrition) are not source of bias	?
c. make sure students not selected for project based on pretest score in Model A	34
VI. Report form completion	
a. include information about testing dates	99
b. follow instructions	26
c. complete all items	29
d. make writing legible	7

Note: The format of the LEA annual evaluation report differs widely from State to State, ranging from the detailed reporting of raw data to supplying only a single overall measure of impact. Thus one SEA with access to raw data was able to check the data closely (and found that about 93% of the reports contained at least one table look-up error) but another State had no way to check because raw data resided only with the LEA's. Therefore not every possible problem could be checked by the USOE team and our estimates of various error rates shown in Table 3.1 are undoubtedly low. For instance, the eleven cases with data based on the use of inappropriate norms tables (see III (a) in the table) are from a subsample of 40 projects which reported raw data to the State; for those 40, this specific error rate is 28%; we do not know for the other 140 or so reports examined.

Procedural. Most of these apply to the general implementation rules and other administrative areas.

1. Provide a detailed special handbook on the implementation of each model. The handbook should be very elementary, in step-by-step flow chart fashion, with plenty of concrete examples of documents and approaches which have worked.

2. Emphasize reduced testing requirements with the proposed models. Too many school personnel are too worried about too many tests. Specify, for each model, the minimum testing possible.

3. To encourage the proper administration of tests, encourage those districts using once-a-year spring testing to have the teacher from the next higher grade give the test. To the third grade teacher testing the second grade students at the end of the year, accuracy would be paramount since next year he would have those very students and would, supposedly, welcome accurate test scores in their folders.

4. Give more guidance regarding test selection. Many studies have demonstrated the importance of test content for detecting student growth in specific skill areas.

5. Add more information to the handbook on out-of-level testing. Most individuals still feel very uncomfortable attempting to implement functional-level testing although they recognize the necessity with some students. Prepare a detailed checklist which States can give to LEA's showing the effort involved in ordering, the logistics of testing, and the scoring and translating of scores. Discuss the pay-offs for these extra steps in terms of an increase in accuracy of the scores. Also, some myths are common and should be investigated and discussed. ("Functional-level testing will result in the pretest average being lower;" "it will result in choosing the wrong students;" "it will give inaccurate estimates of gain;" "it will compare students at one grade unfairly with those at the next lower grade," etc.)

6. Provide guidelines on what action to take if the raw score leads to a converted score too low to be included in the percentile table. For example, a fourth grader might take an out-of-level Metropolitan Achievement Test, Total Reading. The raw score could correspond to a standardized score of 30, but that is too low to be included in the percentile table.

7. Communicate the results of the data collected in this study to publishers, especially the information about the needs for norming earlier in the year, and for less confusing norms tables.

8. Remove the suggestion in current documentation that two-thirds of the project should take place between the two pre-and post-tests. The incidence of failure to follow this requirement is negligible, and should diminish to zero as districts move to appropriate testing dates. The requirement leads to reporting of non-informative data, and to problems when schools have a provision for students to return to their regular classrooms when they have mastered a certain body of material. (Only 2.2% of the projects sampled did not meet this requirement, and they all planned to remedy the problem in the next school year.)

9. Investigate the conditions under which combining results from different grades is appropriate. If a comparison group for Model C is too small within a particular grade, some addition of non-Title I students from the next higher grade might be possible although technically they are not in the project. Or in Model A, adding the two Title I students in 8th grade to those in 7th might reduce the trauma attendant to finding an average of a 10 NCE loss (due to small sample size and unstable data).

10. In the light of the high rate of errors in table reading, consider the use of a raw score reporting system. In the absence of mechanical aids to table look-up, provide simpler score conversion tables to users and communicate the problem to test publishers. (Some publishers have already provided vastly improved tables for those States where the evaluator has insisted upon them.)

11. Provide assistance, both in guidelines and in the computer program, for the plotting of pretest against posttest for all models. The visual inspection of the scatterplots could be expected to indicate floor and ceiling effects, gross non-linearity of relationships, and the presence of unexpectedly high or low "gainers."

12. Classify evaluation errors by severity so that when two implementation suggestions conflict, the Title I person can have some guidance.

13. Many of the details of the proposed system are best communicated in materials to supplement a workshop. The technical reports are excellent, but difficult for SEA's and LEA's to obtain (budget problems, communication concerning new reports and methods of obtaining them, etc). Make them available to SEA's at no cost; SEA personnel could distribute them giving implicit sanction and perserving their leadership.

14. Show some examples of how program objectives could be stated under the new reporting format.

One State requests that the LEA Title I director estimate the NCE gain to be achieved for each grade within each project. If the third grade reading project at Memorial Elementary has a reputation as the best, then a seven NCE gain may be the objective. If it is the worst, then one-half an NCE might be appropriate.

15. Prepare examples to help SEA's and LEA's communicate evaluation results to their various audiences, including use of NCE's, percentiles, etc.

16. Consider alternative methods of tying a cost figure to a project. Since districts generally spend about 75% of their budget for instructional personnel direct costs, reporting only those costs may increase the accuracy and decrease the reporting burden. Similarly, there may be a way to rely merely on GEPA Section 437 data.

Clerical. Clerical suggestions refer to recording, translating, and calculating processes.

1. Instruct LEA evaluators, when looking up average standardized scores in publishers' norms tables, to use the individual percentile norms table, not the school norms table.

2. Remove any requirements for data point interpolation at the district level. Interpolation appears to be more error-prone than it is worth.

3. Send copies of an exemplary testing-dates chart to all interested parties. For example, the forms used in one State include a chart such as the one shown in Appendix 3 which assists district personnel in avoiding test administration date errors.

4. Add a requirement for the project report to provide the average selection NCE, where possible, and the average pretest NCE. This will allow an easy edit check to see if the selection was based completely on the pretest, and if the most needy students were chosen.

5. Revise the percentile-to-NCE conversion tables so that they will be easier to use. Perhaps they could be placed in groups of ten, with clear lines to demarcate the columns. They are error-producing at present.

6. Develop "opti-scan" forms and software for the analysis of scores from major tests. If a State, or large LEA, decides to centralize the scoring process rather than have the local classroom teachers score their own papers, this is the solution with the greatest long-range potential, though perhaps the highest initial expense. Furthermore, though the accuracy and speed of data processing will be vastly improved, the system must also have a set of appropriate error-detection methods built in.

7. Encourage LEA's to score some tests by hand, even if they have engaged a scoring service, in order to check the accuracy of the service.

8. Provide guidance regarding the use of automated data processing as often as possible for the various score conversions and manipulations necessary in the system. Encourage other approaches, too, to preserve the integrity of the data:

(a) Staff should try to perform score conversion activities in teams, with people double-checking the work of others whenever possible;

(b) Raw data should be stored (or sent to the LEA or SEA) to enable later checking of a sample of them regarding the correctness of the tables used, the score conversions, etc.;

(c) Tables should be re-formatted at the local level to allow for easier reading (e.g., 2-column tables are much easier to work with than multiple-column tables); and,

(d) Staff responsible for performing the score conversions should be trained in the use and interpretation of standardized tests, so that the scores that they work with will come to have intrinsic meaning leading to detection of clearly wrong, inappropriate, or out-of-range scores.

Some sort of pre-programmed calculator or computer can also facilitate the data analysis. For small LEA's or SEAs without access to larger computer facilities, a set of programs and implementation materials are being developed for use with hand-held programmable calculators available from $89 and up. (Resource Development Institute, the Technical Assistance Center serving Region VI, has developed these programs, which can be furnished on magnetic cards for at least two machines, the Texas Instruments Programmable 59 and the Hewlett-Packard 65.)

Analytical. The suggestions in this section refer to technical characteristics of the models.

1. Clarify the severity of the regression hazard when two-stage selection takes place. For instance, a potential Title I treatment group of 100 students may be identified by teacher referral. Then a pretest is given to those 100, and the 95 students most in need are given treatment. Obviously the regression effect is considerably less than if the five students most in need were chosen.

2. Investigate further the trade-off in comparing project effects from spring to spring versus fall to spring. Once-a-year testing is much easier on the budget and school time, but may be offset by the loss of subjects and their loss of knowledge over the summer.

3. Assist evaluators in the objective identification of "outliers" (student data so extreme that they are likely in error). For instance, in plotting a Model C implementation, one district evaluator was surprised to find two comparison group individuals who were at the top of the distribution on the pretest and almost at the bottom on the posttest. Examination of their scores revealed that they had scored at the 99th percentile on the reading

comprehension pretest, without missing a single item. In contrast, their vocabulary scores were at the 1st percentile. Clearly some error had been made (by the scoring service, perhaps), and the individuals were dropped from the analysis. The result was that the estimate of project effects changed from negative to positive.

Three conclusions appear evident from Chapter III: (1) The process to change LEA evaluation activities to conform with those prescribed by the evaluation models is laborious and, by necessity, iterative. States we visited had staff pursuing this goal for as long as two years, and many reported that more work is still needed.

(2) Probably the most pervasive administrative problem for SEA's is that it is almost impossible for them to know what actually happens in the LEA's. Of course, this is much more than just an evaluation problem, but it greatly affects evaluation data. The tests may have been given on other than the reported dates (in one of the States, the test was supposedly administered on a Sunday); the test administrator may have ignored the directions for proper test giving; tests designed for group administration may have been given individually or with relaxed time limits; they may have been administered to the wrong individuals; or other situations may have prevailed and no one will ever know.

(3) Another threat to the validity of data reported appears to be the reliance on information that is scored, transformed and analyzed manually. Though conversion to automatic data processing wherever possible in the evaluation system does not promise to be a total panacea, it seems a promising first step.

CHAPTER IV. REGULATIONS

This report has discussed the experiences of SEA and LEA personnel who used the evaluation models during school year 1976-77 (and in a few cases, earlier). All State staff have been encouraged to get the LEA adoption process underway, have attended national workshops regarding the models, and have called on their Technical Assistance Center consultants for help; USOE's role in this has been one of encouragement, persuasiveness (usually), and assistance. We anticipate that those behaviors will continue even as use of the uniform evaluation models becomes required, not merely voluntary. The purpose of this chapter is to discuss our work underway to issue such a requirement in the form of regulations.

In response to Congressional inquiry in early 1975, we sought a legal opinion regarding the need for regulations and were advised that they are required for Section 151. The following paragraphs describe (1) our efforts to solicit input from interest groups and the public, (2) the probable content of regulations, and (3) the anticipated publication procedure and schedule.

Public Participation and Consultation. As indicated in our earlier reports about the model development process we have solicited the advice and suggestions of SEA and LEA representatives regarding appropriate models for Title I evaluation. We will continue our efforts to consult with the public during the regulations process.

In the fall of 1976, in anticipation of the preparation of regulations, a series of informal meetings was held with representatives of the Committee on Evaluation and Information Systems (CEIS) of the Council of Chief State School Officers and with State Title I Coordinators. In addition, we have been contacting representatives of each of the major interest groups in education to inform them of our work to date, and to advise them that regulations are forthcoming. The regulations process will also include an opportunity for public comment after both a memorandum about our intent and Proposed Regulations are distributed.

Regulation Content. Section 151(b) requires the Commissioner to publish "standards of program or project effectiveness." In keeping with this requirement, the regulations will include a set of general technical standards appropriate for Title I evaluations. These standards will be concerned with such issues as the use of comparison groups, the conditions under which sampling is appropriate, the reliability and validity of tests, etc. The establishment of general standards is somewhat problematic because their appropriate application to a specific Title I project may vary depending on the characteristics of that project. In trying to develop the best formulation of technical standards, we are consulting the work of the Joint Committee on Standards for Educational Evaluation (a joint committee representing the major professional organizations in education), and are also doing further empirical work to determine the degree of specification that the standards must contain in order to ensure that the data obtained will be technically sound. This work is concerned with such questions as how reliable an instrument must be in order for an evaluator to have confidence in the data it yields, etc.

The three models that are currently being disseminated describe procedures which, when correctly followed, will yield data that are technically sound. The regulations will describe the major requirements of these three models. We anticipate that most LEA's who are providing instructional services in reading and math in Grades 2 through 12 will use one of the three models. However, an LEA (or an SEA) might wish to develop an alternative to them. Such alternatives would be permitted, provided that they are technically sound and meet the requirements of the legislation. To accommodate the development of acceptable alternatives, the regulations will specify a process for their approval. We should note that these comments reflect our current thinking about regulations; that thinking may certainly be modified as a consequence of suggestions received from the public during the regulations process.

The Regulations Process and Schedule. Our Regulations process will involve (1) the distribution of a memorandum to all States which solicits comments about specific issues regarding upcoming regulations; (2) a Notice of Proposed Rule-making, and (3) Final Regulations. Delays in the many clearance steps required for the publication of regulations have forced us to revise our publication schedule, and earlier plans for a formal Notice of Intent were not approved by the Office of the Secretary.

We hope to publish Proposed Rules in the fall of 1978, and Final Regulations during the following winter or spring. This means that regulations governing Section 151 would be in effect during the school year 1979-80. However, as indicated in Chapter III, many States will go through several steps in implementing the new regulations, and there will be considerable variation in the speed with which different States achieve implementation. Those who are already trying the models on a voluntary basis should be able to report data from the school year 1979-80 in their November 1980 report to the Commissioner. Other States will require a longer implementation period. This means that USOE's receipt of uniform data from local and state reporting will probably increase gradually through the next few years. We have work underway now--which we will report on by the end of the summer, 1978--to estimate the amount of aggregatable data to be received each year and the degree to which national estimates may result from them (with assorted accompanying figures to estimate the effects of missing data, etc.).

CHAPTER V. CONCLUSIONS

The three substantive chapters of this report described (a) the current status of the use of new Title I evaluation models nationwide; (b) the experiences of State and local personnel who have tried the procedures and their reactions and (c) the process underway simultaneously in USOE to draft and publish regulations in this area. This documentation of the evaluation program in terms of various activities and some of their effects also suggests some further work to facilitate LEA and SEA evaluations of Title I.

The purpose of this chapter is to highlight some of our conclusions about the work at this time and to describe briefly some of the activities planned for the future which are implied by those conclusions. The chapter is organized to address four major topics: (1) services to support the use of Title I evaluation models, (2) documents to further describe the models, (3) focused investigations on remaining questions, and (4) solicitation and use of public input.

Conclusions and plans regarding assistance service

Discussions with SEA and LEA personnel have continually emphasized the importance of on-site, specific advice regarding evaluation activities. Often principles can be stated and general advice offered in handbooks, training sessions, etc., but unique characteristics of projects, participants, and/or setting demand that special consideration come into play for the application of those principles. Personnel in the Technical Assistance Centers play a vital role in this area, and current indications are that they fulfill the role extremely well.

Now that the Centers have been in operation for nearly two years, we feel this is an appropriate time for a third-party assessment of their work and function. A contract to support such an effort is underway as well as continuing discussions with State and local personnel.

SEA's have been surveyed to determine their anticipated need for services during the period ending March 1, 1979, and the Centers have been funded to meet that need. (There are also provisions for meeting increased needs if estimates prove to be too low). Federal regulations dictate that we conduct a competitive procurement to award the contracts for provision of these services after March of 1979. Our intention is to insure, however, that services not be interrupted during such a procurement. Other services as provided by USOE personnel directly or as augmented by additional contracts (for example, for the provision of materials and assistance with national workshops to introduce SEA personnel to evaluation models for the Title I program for the neglected or delinquent) will also continue.

Documents about the models and their procedures

USOE policy has been that the Government Printing Office would be the primary agent for distributing handbooks, pamphlets, and other materials about the proper conduct of Title I evaluations. It was felt that printing and distribution of perhaps 20,000 copies of documents would be impossible for USOE to accomplish.

Difficulties faced by SEA's and LEA's in obtaining our materials have made us realize that our early decision (to rely primarily on GPO and only supplement those efforts) was wrong. We now envision a more active role for USOE in the provision of materials--contacting States to determine the number of copies needed, having a sufficient number printed, and mailing them from USOE. We will also maintain a stock of USOE materials in the Technical Assistance Centers for their personnel to distribute. The third source of materials will be GPO to enable persons who do not interact with States or the Centers to obtain copies.

Another effort growing from SEA and LEA requests will be computer software and documentation which is more generalizable than our current version. It will include provsions for entering data in various formats, for performing edit checks on the data, for printing out scatter-plots, etc. We hope this will improve the accuracy in computation, etc., that is now a problem. Work in this will begin under contract in fiscal year 78, and copies of the program and documentation will be distributed widely as described above.

Persistent confusion regarding specific decisions in the conduct of an evaluation as well as recent recommendations by the Lawyer's Committee for Civil Rights under Law under contract to NIE (Silverstein and Schember, 1977) have identified the need for a technical policy manual in the area of Title I evaluations. Our plans are that it will be organized topically to facilitate referencing it and will include many examples. It will provide guidance about specific practices only implied generally by the regulations, and we hope to have a draft available for review by interested parties simultaneously with their review of draft rules.

Focused investigations to answer remaining questions

Problems still arise in districts regarding decisions about specific evaluation practices. Many of those decisions involve trade-offs about which there is currently no information. For example, some investigators have noted inaccuracies in scales across test levels developed to allow the use of a lower or "functional level" test with students and still describe their performance with respect to that of their age-level peers. Others note that there are inaccuracies, too, when students are given tests too difficult for their developmental level. Local evaluators, then, must judge which type of error will be worse (that due to the scaling problems or that due to a mismatch between items and student level), and they need more information by which to make such judgments. We plan to re-analyze existing data bases (from completed Federal evaluation studies) for evidence regarding such trade-offs, and that evidence will be much of the content of the technical policy manual described above.

Work is also underway regarding evaluation models for other parts of the Title I program, such as that providing services for migrant children, for the neglected or delinquent, and for pre-schoolers. Each is being addressed in a separate contract, and USOE, State, and LEA personnel serve on advisory panels to react to feasibility and validity issues as the work proceeds.

A third area of current investigation is focusing on the amounts of usable LEA and SEA data USOE will have at different times for reporting to Congress. We plan to estimate these amounts and develop means for using them by next fall. (All that time we will also have more definite information about the timing of regulations, allowing us also to predict when we will have uniform, comparable data from all SEA's.)

The importance of continued public input

Section 151 emphasizes SEA and LEA involvement in this work to develop evaluation models, and the program to implement the Section has relied heavily on their input. SEA and LEA personnel are on panels focusing on model development activities for each of the parts of Title I; additional committees have been formed to assist with the drafting of regulations;

There is much interaction among personnel from USOE (both in Title I and in evaluation) SEA's (again representing both the program and evaluation areas), LEA's, professional groups, and commercial concerns (for example, test publishers). Seldom is there unanimity across these groups regarding policy for Title I evaluations, but the participation of people with varying concerns can at least keep USOE recommendations from being more restrictive than necessary.

In short, it is rewarding to see changes which mark improvement in the state-of-the art of educational evaluation. One can cite, for example, changes in the types of scores and services offered by test publishers; the increase in the number of textbooks on evaluation topics (often including the same topics as those emphasized in the Title I materials); more discussion at professional meetings; critical examination and use of evaluation data for identification of effective practices or for changing programs; and increased interest on the part of parents. While it is not appropriate to assume that the Title I evaluation system has been solely responsible for this trend, we do feel it has played an important role in it and has been, in certain instances, a catalyst both for the improvement of evaluation practices and for the further examination of related issues.

☆U.S. GOVERNMENT PRINTING OFFICE: 1978-260-934/2077